Business Patterns Data Set: Technical Documentation

November 2006

Beata Bajaj and Kathryn L.S. Pettit
The Urban Institute

The Fannie Mae Foundation creates affordable homeownership and housing opportunities through innovative partnerships and initiatives that build healthy, vibrant communities across the United States. The Foundation is specially committed to improving the quality of life for the people of its hometown, Washington, D.C., and to enhancing the livability of the city's neighborhoods. The Foundation, a private nonprofit organization supported solely by Fannie Mae, has regional offices in Chicago, Dallas, and Pasadena, Calif.

The Fannie Mae Foundation, through its publications and other programs, is committed to the full and fair exposition of issues related to affordable housing and community development. The opinions expressed in this publication are those of the author(s) and do not necessarily represent the views of the editors, the Fannie Mae Foundation, or its officers or directors.

The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders.

Introduction

In addition to a wide selection of demographic, housing, and mortgage lending data, the Fannie Mae Foundation's DataPlace™ Web site includes key economic indicators from the U.S. Census Bureau's Business Patterns data series. The indicators currently are available from 1998 to 2003 and will be updated annually. DataPlace™'s Business Patterns data are available for ZIP codes, cities, counties, and larger geographic entities, enabling users to learn about the changing spatial patterns of economic activity in their communities and regions.

This document describes the Business Patterns series source data set and the summary indicators from this set that are included in DataPlace™. It also discusses technical issues associated with preparing the Business Patterns data for DataPlace™. A companion analytic brief, *Business Patterns and Trends: National Summary*, is available at http://www.dataplace.org/guides.html.

Description of Business Patterns Source Data Set[1]

The U.S. Census Bureau's Business Patterns series is produced annually and provides subnational economic data by industry. The series is useful for studying the economic activity of small areas; analyzing economic changes over time; and benchmarking statistical series, surveys, and databases between economic censuses.

The Business Patterns series provides information on number of establishments, employment, and payroll. (DataPlace™ includes data based on only the first two of these categories.) An establishment is a single physical location at which business is conducted or services or industrial operations are performed. It is not necessarily identical with a company or enterprise, which may consist of one or more establishments. When two or more activities are carried on at a single location under a single ownership, all activities are generally grouped together as a single establishment. The entire establishment is classified on the basis of its major activity and all data are included in that classification. Establishment counts in the Business Patterns series represent the number of locations with paid employees at any time during the year.

The Business Patterns series also provides data on paid employment for full- and part-time employees, including salaried officers and corporate executives on payroll in the pay period including March 12. Employees on paid sick leave, holidays, and vacations are included; proprietors and partners of unincorporated businesses are excluded. Self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees also are excluded.

[1] Descriptions of the Business Patterns series, data content, and industrial sector coverage are based closely on U.S. Census Bureau, *County Business Patterns: United States: 2002*, CBP/02-1, issued November 2004.

Data Collection

The Census Bureau extracts the Business Patterns data from the Business Register, a file of all known single- and multi-establishment companies. The Annual Company Organization Survey and quinquennial Economic Censuses provide individual establishment data for multilocation firms. Data for single-location firms are obtained from various programs conducted by the Census Bureau, such as the Economic Censuses, the Annual Survey of Manufactures, and Current Business Surveys, as well as from administrative records of the Internal Revenue Service, the Social Security Administration, and the Bureau of Labor Statistics.

Industry Sector Classification and Coverage

Industry sector classifications of Business Patterns data currently are based primarily on the North American Industry Classification System (NAICS).[2] The Business Patterns series includes the NAICS sectors listed in Table 1. (Groupings of NAICS sectors used in DataPlace™ are described below under "Business Patterns Data in DataPlace™".)

Table 1. NAICS Sectors Included in Business Patterns Data

Sector	Description
11	Forestry, Fishing, Hunting, and Agriculture Support
21	Mining
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
95	Auxiliaries (except corporate, subsidiary, and regional management)
99	Unclassified

The Business Patterns data exclude the following NAICS industries: crop and animal production (NAICS 111, 112); rail transportation (NAICS 482); National Postal Service (NAICS 491); pension, health, welfare, and vacation funds (NAICS 525110, 525120, 525190); trusts, estates, and agency accounts (NAICS 525920); private households (NAICS 814); and public administration (NAICS 92). The Business Patterns data also exclude governmental establishments, except for wholesale liquor establishments (NAICS 4228), retail liquor stores

[2] NAICS, the official industrial classification system of the United States, is used to group establishments that apply the same or similar processes to produce goods or services.

(NAICS 44531), federally chartered savings institutions (NAICS 522120), federally chartered credit unions (NAICS 522130), and hospitals (NAICS 622).

Additional descriptions of NAICS sectors are provided in the manual *North American Industry Classification System: United States, 1997.* Information on ordering this manual is available at http://www.census.gov/epcd/www/naics.html.

Before 1998, the Business Patterns series was based on the Standard Industrial Classification (SIC) System. Although the NAICS and SIC classification systems can be bridged using cross-reference files provided by the Census Bureau, they are not entirely comparable because not all SIC categories have one-to-one correspondence with NAICS categories. Because limited comparability could create inaccuracies in our tabulations of the Business Patterns data, DataPlace™ does not include any data preceding the introduction of the NAICS codes in 1998.

Employment Coverage

The Business Patterns data cover most of the nation's economy, but as noted above they do not represent all industrial sectors or all types of economic activity. A comparison with employment data reported in the Bureau of Labor Statistics Current Employment Statistics (CES) program is illustrative, as the CES provides one of the broadest measures of employment in the United States. As of March 2002, the CES reported U.S. employment of 129.7 million. For the same period, the Business Patterns series reported total employment of 112.4 million, 87 percent of the CES figure.

Coverage of the Business Patterns data can vary substantially from place to place, depending on the composition of the local industrial base. Coverage is likely to be particularly low in areas with large numbers of government employees, who for the most part are not included in Business Patterns data. An extreme example of such an area is the District of Columbia, where Business Patterns data reported employment of 419,000 in 2002, 63 percent of District employment recorded in the CES for the same period.

Business Patterns Data in DataPlace™

DataPlace™ Indicators

DataPlace™ uses the Census Bureau's Business Patterns series to develop a set of establishment and employment indicators (see Table 2). The site provides counts, per capita rates, and percent distributions of establishments and employment by industry and establishment size category. It is important to note that indicators of employment by industry are created only for counties and larger geographies because of a lack of industry-specific employment data at the ZIP code level in the Census Bureau source data files.

Table 2. Business Patterns Indicators in DataPlace™

Total number of establishments
Establishments per 1,000 population
Total employment
Employment per 1,000 population
Number of establishments by establishment size category
Percent of establishments by establishment size category
Establishments by establishment size category per 1,000 population
Employment by establishment size category
Percent of employment by establishment size category
Employment by establishment size category per 1,000 population
Number of establishments by general industry classifications
Percent of establishments by general industry classifications
Establishments by general industry classifications per 1,000 population
Employment by general industry classifications (for counties and larger areas)
Percent of employment by general industry classifications (for counties and larger areas)
Employment by general industry classifications per 1,000 population (for counties and larger areas)
Number of establishments by detailed industry classifications
Percent of establishments by detailed industry classifications
Employment by detailed industry classifications
Percent employment by detailed industry classifications
Number of establishments for selected residential services sectors
Establishments for selected residential services sectors per 1,000 population

Establishment Size Categories and Industry Classifications

For the indicators based on establishment size, DataPlace™ uses nine categories: 1-4 employees, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, and 1,000 or more employees.

Three typologies stratify the industry-specific data on establishments and employment (see Table 3). The first two industry typologies, one general and the other more detailed, aggregate the source data using two-digit NAICS industry codes. The third typology, which describes residential service sectors, relies on much more detailed industry categories (four- to six-digit NAICS codes).

Table 3. DataPlace™ Industry Typologies

General Typology		
Number	**Category Description**	**NAICS Codes Included**
1	Agriculture, Forestry, Mining	11, 21
2	Manufacturing	31-33
3	Utilities; Construction; Transportation	22, 23, 48-49
4	Wholesale Trade; Retail Trade	42, 44-45
5	Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; and Administrative and Support and Waste Management and Remediation	51-56
6	Other Services: Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; Accommodation and Food Services; Other Services; Auxiliaries; Unclassified	61, 62, 71, 72, 81, 95, 99
Detailed Typology		
Number	**Category Description**	**NAICS Codes Included**
1	Agriculture, Forestry, Fishing, Hunting	11
2	Mining	21
3	Manufacturing	31-33
4	Utilities	22
5	Construction	23
6	Transportation and Warehousing	48-49
7	Wholesale trade	42
8	Retail trade	44-45
9	Information	51
10	Finance and Insurance	52
11	Real Estate, Rental and Leasing	53
12	Professional, Scientific, Technical Services	54
13	Management of Companies, Enterprises	55
14	Administrative, Support, Waste Management Services	56
15	Educational Services	61
16	Health Care and Social Assistance	62
17	Arts, Entertainment, and Recreation	71
18	Accommodation and Food Services	72
19	Other Services (except Public Administration)	81, 95, 99
Residential Services Typology		
Number	**Category Description**	**NAICS Codes Included**
1	Groceries/supermarkets	445110
2	Convenience stores	445120
3	Drugstores	446110
4	Commercial banks	522110
5	Offices of physicians	all starting with 6211
6	Offices of dentists	all starting with 6212
7	Social assistance services	all starting with 6241
8	Full-service restaurants	722110
9	Other food services (cafeterias, snack bars, mobile food services)	all starting with 72221 and 722330
10	Personal care services (barber shops, beauty salons)	all starting with 8121
11	Dry cleaners/laundries	all starting with 8123

Geographic Levels Included in DataPlace™

DataPlace™ Business Patterns indicators are created for eight geographies: ZIP Code Tabulation Areas (ZCTAs — see Technical Issues section for details), places, counties, states, metropolitan areas, regions, divisions, and the United States.

The Census Bureau's County Business Patterns (CBP) data files were used to create indicators for counties, states, divisions, regions and the nation. In addition to county-level data, the CBP

source files include separate summaries for states and the United States; therefore, the county-level data did not have to be aggregated to obtain data at these geographic levels.

The Census Bureau's ZIP Business Patterns (ZBP) data files are the source of DataPlace™'s ZIP code–level data. (The Technical Issues section explains in detail how the ZBP data were assigned to ZCTAs.) To create the Business Patterns data for places, the ZCTA-level ZBP data were weighted using a population-based weighting file downloaded from the Mable/GeoCorr2k Web site and then aggregated to the place level.[3] Only places that are central cities or have population over 50,000 are included in DataPlace™.

We used the Office of Management and Budget's (OMB) 1999 metropolitan area definitions when assembling the Business Patterns data for metropolitan areas.[4] We added up the county-level CBP data for all metropolitan areas outside of New England. In New England, where metropolitan areas follow municipal rather than county boundaries, we used a different approach. First, we used GIS software to identify ZIP codes in the New England metropolitan areas. We then aggregated ZCTA-level ZBP data to create the metropolitan-level data.

[3] The Missouri Census Data Center developed Mable/GeoCorr2k: Geographic Correspondence Engine with Census 2000 Geography. Mable/GeoCorr2k is a free online tool that provides area, housing unit, or population weights for Census 2000 geographies. For more information, visit their site at http://mcdc2.missouri.edu/websas/geocorr2k.html.

[4] In the future, DataPlace™ will include data for metropolitan areas as defined by OMB in 2005.

Technical Issues

This section describes several technical issues related to the processing of Business Patterns data for DataPlace™.

Data Suppression and Exclusion in the Census Bureau Source Files

While CBP and ZBP provide a full count of the number of business establishments by area, the Census Bureau suppresses data on total employment in individual counties, and more frequently ZIP codes, where confidential data may be at risk of disclosure. For example, if a ZIP code has one very large establishment and only a few other small firms, publishing precise data on total employment for the ZIP code could give competitors a fairly good idea of that establishment's employment level.

Additionally, to avoid disclosure of information on individual companies in detailed tabulations, the Census Bureau completely withholds from the source files employment data broken down by industry or establishment size class for *all* ZIP codes. Similarly, the Census Bureau excludes employment data by establishment size class for *all* counties (see Table 4, upper left panel).

For the Business Patterns data records affected by suppression or exclusion, we have developed a method to impute employment, which is described in the next section. The imputation procedure has very little effect on aggregate employment estimates. For example, the sum of employment in all ZIP codes in the 2000 DataPlace™ file (including the imputed values) is less than 1 percent below the separately published CBP national total for that year. However, the imputation procedure could result in errors in total employment estimates for individual counties and ZIP codes affected by data suppression, bound by the range of employment given by the suppression flag.[5] In DataPlace™, between 14 and15 percent of ZIP codes and less than 1 percent of counties in any given year have imputed total employment values.

Imputing Suppressed or Excluded Employment Values

The suppression or exclusion of Business Patterns employment data varies depending on the geography and level of industry and establishment-size detail, as shown in Table 4. The table may be useful in understanding the following explanations of methods for imputing suppressed or excluded employment values.

Table 4. Comparison of Business Patterns Data in Census Bureau Source Files and in DataPlace™

	Source Data (provided by the Census Bureau)	DataPlace™ Data

[5] In some instances, where the Census Bureau's source files do not contain an exact total employment figure for a given geographic area, the suppression flag will provide a range of total employment; for example, 20 to 99 employees. In other instances, the suppression flag will simply indicate whether data has been suppressed, without providing an employment range. The two types of suppression flags are distinguished in Table 4.

	ZIP Code	County	State/US	ZIP Code/ Place/ and NE MSA [2]	County/MSA	State/US
Employment						
Total Employment	Yes, with suppression range flag	Yes, with suppression range flag	Yes, with suppression range flag	Yes, with suppressed values imputed	Yes, with suppressed values imputed	Yes
Total Empl. for 2-digit Industry	No	Yes, with suppression range flag	Yes, with suppression range flag	No	Yes, with suppressed values imputed	Yes, with suppressed values imputed [3]
Total Empl. for 3-6 digit Industry	No	Yes, with suppression range flag	Yes, with suppression range flag	No	No	No
Total Empl. by Estab. Size Category [1]	No	No	Yes, with yes/no suppr. flag	All values imputed by UI	All values imputed by UI	Yes, with suppressed values imputed
Establishments						
Total Establishments	Yes	Yes	Yes	Yes	Yes	Yes
Total Estab. for 2-digit Industry	Yes	Yes	Yes	Yes	Yes	Yes
Total Estab. for 3-6 digit Industry	Yes	Yes	Yes	Resident Services Only	Resident Services Only	Resident Services Only
Total Estab. by Estab. Size Category [1]	Yes	Yes	Yes	Yes	Yes	Yes

Notes:

1. Although the Census Bureau's source files break industry-specific employment and establishment totals out by establishment size categories (for specified geographies only), we do not create indicators of employment or establishments by industry by establishment size.
2. DataPlace™ uses aggregated ZIP code–level data to construct indicators for places and the metropolitan areas in New England.
3. None of the 21 two-digit industry sectors within the United States had suppressed numbers for total employment, and as a result no employment values had to be imputed for the industry sectors at the national level. However, some imputation of total employment data was necessary for the two-digit industry sectors for states.

Imputation of employment data is made possible by the fact that the Census Bureau reports establishment counts by establishment size category for all Business Patterns data, regardless of geography or level of industry detail. We take advantage of this information to impute all suppressed and excluded employment figures, with the exception of employment by industry for all ZIP codes and employment within residential service sectors for all geographies.

We first imputed total employment within each of the nine establishment size categories by multiplying the number of establishments within the category by the national average number of employees for establishments of that particular size class.[6] For example, in 2001 establishments with 1-4 employees had an average of 1.7 employees, and those with 5-9 employees had an average of 6.6 employees. These averages were calculated using printed U.S. County Business Patterns reports available on the Census Bureau Web site: http://www.census.gov/prod/www/abs/cbptotal.html.

[6] At the county and ZIP code levels, it is necessary to impute employment for all nine establishment size categories, because the Census Bureau excludes these figures from the source files. However, at the national and state levels, the Census Bureau reports employment by size category, so we only impute any suppressed values.

Once these employment data were imputed, as needed, employment was summed across individual establishment size categories. If the Census Bureau suppressed the total employment figure, we used this sum in its place, after adjusting it to correspond to the range of values specified by the suppression flag. Where these range adjustments were needed, the imputed employment data for individual establishment size categories were modified accordingly. If, on the other hand, the Census Bureau did not suppress the total employment figure, this sum was compared with the reported total number of employees for a given geographic area or for a given industry sector within a geographic area. If we found any discrepancies between the sum and the reported figure, the imputed employment data for individual establishment size categories were in turn adjusted to add up exactly to the total employment figure reported by the Census Bureau.

We imputed indicator series only where a suppression flag provided a range of values for the suppressed total employment figure. This valid range allowed us to verify our imputed values. The Census Bureau's Business Patterns series report, or provide suppression flags for, employment data by industry for counties, states, and the United States, but not for ZIP codes. For that reason, DataPlace™ includes employment data by industry for counties but does not include these data for ZIP codes. The methods described above could be used to impute industry employment totals for ZIP codes, but the process of reconciling estimated employment by industry with the total employment would produce many inconsistencies.

Assigning Geographic Boundaries to ZIP Codes

ZIP codes are merely postal routes (i.e., collections of addresses) and therefore do not have official geographic boundaries comparable to those for other recognized territorial subdivisions in the United States (e.g., census tracts or congressional districts). Because ZIP codes are used to tabulate data from many sources, the Census Bureau drew geographic boundaries for them nationwide in 2000, creating a set of ZIP Code Tabulation Areas (ZCTAs)[7]. These polygons best approximated the spatial extent of ZIP code postal routes at that time.

Because ZCTAs are the base geographic boundary file for DataPlace™ and because we associate the Business Patterns data with Census 2000 ZCTA data when we calculate per capita rates, we matched ZIP codes from the Business Patterns files with ZCTAs. To accomplish this, we used three files: the Census Bureau's ZBP file, the Census Bureau's ZCTA 2000 file, and a commercial ZIP code boundary file from Bamberg-Handley, Inc. First, we matched the ZBP five-digit ZIP codes with the ZCTA codes. Direct matches occurred in a substantial majority of cases (about 80 percent). We did not attempt to evaluate or adjust for any geographic differences between the ZCTA and ZIP code boundaries.

We then used mapping software to try to identify the locations of those ZBP ZIP codes that did not match the ZCTA codes. These ZIP codes typically encompass very small areas with

[7] For more information about ZCTAs, see the Census Bureau Web site at http://www.census.gov/geo/ZCTA/zcta.html

particularly intense postal activity. They may be assigned to individual buildings (e.g., the headquarters of a federal agency) or building complexes (e.g., a university). We used the commercial ZIP code boundary file from Bamberg-Handley, Inc., to compare these ZIP code point locations with ZCTA polygons. In 4 percent of these cases, we successfully located ZIP codes within the physical boundaries of a ZCTA and were able to add the relevant ZBP data to the data already assigned to the surrounding ZCTA.

ZBP ZIP Codes Without ZCTA Matches

For the remaining 96 percent of nonmatching ZBP ZIP codes (i.e., 17 percent of all ZIP codes with ZBP data), we were unable to make assignments using mapping software. Fortunately, these unassigned ZIP codes tend to have small employment levels. Three-quarters of them contain fewer than 40 jobs, and in total they account for only 4 percent of all establishments and 2 percent of all employment. Because population data are not available for these ZBP ZIP codes without ZCTA matches, no per capita indicators could be computed for these areas. These small ZIP code areas are not likely to be of great interest to community development analysts. Nonetheless, it is important to note that because DataPlace™ uses ZCTA boundaries as the master file for ZIP code–level data, these unmatched ZIP codes will not display on any map or table on the site. DataPlace™ will report “Not Available” should users attempt to access data for these particular ZIP codes.

Table 5. Match Rates between ZBP ZIP Codes and ZCTAs

Matching Status		Year					
		1998	**1999**	**2000**	**2001**	**2002**	**2003**
ZBP ZIP codes that match with the ZCTA 2000 file	**% of ZIP codes**	83.5	82.9	82.6	82.7	83.5	83.5
	% of Total Establishments	97.3	97.3	97.0	96.9	97.6	97.7
	% of Total Employment	95.6	95.8	95.5	95.2	95.5	95.5
ZBP ZIP codes that do not match with the ZCTA 2000 file	**% of ZIP codes**	16.5	17.1	17.4	17.3	16.5	16.5
	% of Total Establishments	2.7	2.7	3.0	3.1	2.4	2.3
	% of Total Employment	4.4	4.2	4.5	4.8	4.5	4.5

ZCTAs that did not match any of the ZBP ZIP codes are considered to be valid geographic areas that simply have no business establishments or employment. Therefore, these ZCTAs are part of the Business Patterns files posted on the DataPlace™ site but have establishment and employment figures set to zero.

Changing ZIP Code Boundaries and Their Relation to Other Geographic Areas

An additional complexity is that the Postal Service frequently adds, deletes, and modifies ZIP codes. For example, when we add up establishments and employment for a particular metropolitan area, the list of ZIP codes changes modestly from year to year, even though we are likely to be dealing with the same basic territory. The percentage of additions and deletions

is small: 95 percent of the ZIP codes (representing more than 99 percent of establishments and employment) reported data for all years from 1997 to 2003.

We did not attempt to identify and reconcile ZIP code boundary changes over time for the DataPlace™ files. A typical change occurs when a large ZIP code in a high-growth suburban fringe area is divided into two or more smaller ZIP codes. In rare cases, ZIP codes can be reassigned to a different location altogether. We expect that areas with rapid population growth, such as outer suburban areas, would have more boundary and assignment changes than core urban areas or areas without much population change. Where such redefinitions have occurred in individual ZIP codes, it would not be accurate to compare changes in ZBP data over time. Any user requiring detailed information on a specific ZIP code area should confirm the history of the area's boundary before making decisions based on DataPlace™'s Business Patterns data.

Also, ZIP code boundaries sometimes overlap county and municipal boundaries. To calculate totals for individual municipalities in DataPlace™, we used a method that involved allocating a share of a ZCTA's employment and establishments to a municipality based on its share of population in the municipality. The method relied on a crosswalk obtained from the Geographic Correspondence Engine from the Missouri Census Data Center (http://mcdc2.missouri.edu/websas/geocorr2k.html).

For the companion DataPlace™ analytic brief mentioned in the beginning of this guide, we needed to assign ZIP codes to metropolitan areas. To do this, we first matched ZIP codes to counties using a 1999 crosswalk produced by the U.S. Census Bureau (http://www.census.gov/geo/www/tiger/zip1999.html). For most of the country, once a ZIP code was assigned to a county, we could then assign it to a metropolitan area, since metropolitan areas are normally defined as aggregates of counties. The exception was New England, where metropolitan areas are defined as aggregates of towns. For New England, mapping software was used to assign ZIP codes to metropolitan areas.

Relating Business Patterns Data to 2000 Census Population Data

Finally, it is important to note that DataPlace™ does not include annual population updates at the ZIP code level. As a result, 2000 Census data are used for all years of the Business Patterns data to create per capita indicators. This limitation may affect the accuracy of per capita indicators in areas with rapidly changing populations.